

# Transmit Signal Design for Optimal Estimation of Correlated MIMO Channels

Jayesh H. Kotecha, *Member, IEEE*, and Akbar M. Sayeed, *Senior Member, IEEE*

**Abstract**—We address optimal estimation of correlated multiple-input multiple-output (MIMO) channels using pilot signals, assuming knowledge of the second-order channel statistics at the transmitter. Assuming a block fading channel model and minimum mean square error (MMSE) estimation at the receiver, we design the transmitted signal to optimize two criteria: MMSE and the conditional mutual information between the MIMO channel and the received signal. Our analysis is based on the recently proposed virtual channel representation, which corresponds to beamforming in fixed virtual directions and exposes the structure and the true degrees of freedom in the correlated channel. However, our design framework is applicable to more general channel models, which include known channel models, such as the transmit and receive correlated model, as special cases. We show that optimal signaling is in a block form, where the block length depends on the signal-to-noise ratio (SNR) as well as the channel correlation matrix. The block signal corresponds to transmitting beams in successive symbol intervals along fixed virtual transmit angles, whose powers are determined by (nonidentical) water filling solutions based on the optimization criteria. Our analysis shows that these water filling solutions identify exactly which virtual transmit angles are *important* for channel estimation. In particular, at low SNR, the block length reduces to one, and all the power is transmitted on the beam corresponding to the strongest transmit angle, whereas at high SNR, the block length has a maximum length equal to the number of active virtual transmit angles, and the power is assigned equally to all active transmit angles. Consequently, from a channel estimation viewpoint, a faster fading rate can be tolerated at low SNRs relative to higher SNRs.

**Index Terms**—Conditional mutual information, correlated MIMO channels, covariance feedback, majorization, minimum mean square error, multiantenna systems, pilot-based channel estimation, Schur-concavity, Schur-convexity, uniform linear array.

## I. INTRODUCTION

**M**ULTIANTENNA communication systems are gaining prominence due to the higher capacity and reliability they can afford [2], [3]. Perfect knowledge of channel state is often assumed in the analysis of such systems. However, in practice, the channel has to be estimated, typically using pilot symbols. Accurate and efficient channel estimation is of critical importance in the design of coherent communication systems. In multiple-input multiple-output (MIMO) channels, this

problem is particularly challenging, due to the large number of channel parameters to be estimated in general. Under the idealized assumption of independent and identically distributed (i.i.d.) channel coefficients, the solution is relatively straightforward due to the i.i.d. nature of the coefficients [4], [5]. However, this idealized assumption does not generally hold in practice, and hence, a study of correlated channels is of interest. In this work, we investigate transmit signal design for optimal estimation of narrowband correlated MIMO Rayleigh fading channels, assuming that the receiver and the transmitter<sup>1</sup> have knowledge of the second-order channel statistics.<sup>2</sup> It is assumed that the receiver uses a minimum mean square error (MMSE) channel estimator. The covariance feedback information is exploited by the transmitter to design the transmit signal to minimize the channel estimation error at the receiver. We assume a block flat fading model, where the channel is constant over a block of transmitted symbols but changes independently from block to block. We design the transmit signal to optimize one of two criteria: minimization of the MMSE at the receiver or maximization of the conditional mutual information (CMI) between the channel and the received signal.

A MIMO channel with  $P$  transmit and  $Q$  receive antennae has a maximum of  $PQ$  unknowns to be estimated. However, correlated MIMO channels possess fewer degrees of freedom, and hence, fewer than  $PQ$  parameters need to be estimated. In view of the large number of channel coefficients to be estimated, we exploit this important fact to design efficient signalling schemes for optimal channel estimation. We consider a general model for correlated MIMO channels that exposes the true degrees of freedom of the channel. Two important channel models are special cases for our general model. The first example is the *virtual channel representation* that was recently proposed in [7] for uniform linear arrays (ULA) at both the transmitter and receiver. The second example is the channel where the correlation in the transmit and receive arrays induce correlation in the rows and columns of the channel matrix [8]–[10] and will be denoted as the *transmit and receive correlated model*. The virtual representation characterizes the channel in the spatial domain by beamforming in the direction of fixed virtual angles determined

<sup>1</sup>This can either be estimated at the transmitter in TDD systems (due to reciprocity of the uplink and downlink channels) or estimated by the receiver and sent to the transmitter via a feedback channel, which is often called covariance feedback.

<sup>2</sup>The rate at which *channel statistics* change depends on large-scale variations in the scattering environment and is typically low. The actual channel coefficients, on the other hand, can vary at a much faster rate due to the phase variations induced by the relative motion between the transmitter/receiver and scattering objects. Thus, in general, channel statistics vary much more slowly and can be estimated more reliably, which allows for the possibility of covariance feedback to the transmitter [6].

Manuscript received November 21, 2002; revised May 3, 2003. This work was supported in part by the National Science Foundation under Grants CCR-9875805 and CCR-0113385 and the Office of Naval Research under Grant N00014-01-1-0825. The associate editor coordinating the review of this paper and approving it for publication was Dr. Helmut Bölcskei.

The authors are with the Department of Electrical and Computer Engineering, University of Wisconsin at Madison, Madison, WI 53706 USA (e-mail: jkotecha@ece.wisc.edu; akbar@engr.wisc.edu).

Digital Object Identifier 10.1109/TSP.2003.821104

by the spatial resolution of the antenna arrays and is analogous to representing the channel in beamspace or wavenumber domain. An important characteristic of the virtual representation is that the nonvanishing virtual coefficients are always approximately uncorrelated and represent the degrees of freedom in the channel and, hence, the essential parameters to be estimated. These degrees of freedom are governed by the scattering geometry, the antenna spacings, and the number of antennae. Channel estimation can now be viewed as the identification of this scattering geometry and estimation of the corresponding nonvanishing channel coefficients. In fact, the nonzero blocks in the virtual matrix indicate the directions in which scattering clusters are present and the couplings between the corresponding virtual transmit and receive angles [7]. Our optimum channel estimation design exploits this structure. The general model considered in this paper exposes the true degrees of freedom in the eigen domain, which corresponds to the virtual domain in ULAs. Hence, estimation of the channel can be equivalently performed in the eigen or virtual domain. We note that in [11], optimal transmit diversity precoder design is discussed for a multiple-input single output (MISO) system to minimize channel estimation error according to the two criteria used in this paper. Hence, the results obtained there are a special case of our design.

We show that the optimal transmit signal is a block signal consisting of beams transmitted in succession along the active virtual transmit angles (or transmit eigen-directions), corresponding to directions in which scatterers are present. We assume that while the total transmit power in a block is constant, the transmit power during each symbol transmission may be different. The power transmitted along the beams is determined by *water filling* solutions [12] resulting from the two criteria. Equivalently, the scattering environment is scanned along the virtual transmit angles one by one to determine the presence of scattering clusters by measuring the signals along the virtual receive angles for each transmitted beam. In other words, the  $i$ th transmitted beam is used at the receiver to determine the  $i$ th column of the channel matrix in the virtual (or eigen) domain. Power is possibly assigned to a beam only if the second-order statistics indicate the presence of significant scattering in that direction. However, the power assigned to the transmit beams depends on the signal-to-noise ratio (SNR) as well. Our optimal signal design suggests that for a given SNR, only those columns of the virtual channel matrix should be estimated that are deemed *dominant* by the water filling criteria. In essence, those channel coefficients whose power is small compared with the background noise are not important from a communication viewpoint, and the transmit power is better utilized in channel coefficients that exhibit a higher SNR. This important observation comes up in a variety of related contexts, including noncoherent communication scenarios, capacity-maximizing water filling solutions in the case of imperfect channel state information [6], [13], [23], the optimality of beamforming at low SNRs [6], and space-time coding for correlated channels [14], all of which indicate the importance of such dominant degrees of freedom.

Section II introduces the MIMO channel model. Section III discusses MMSE estimation of the MIMO channel and motivates criteria for optimum signal design. In Section IV, we

obtain the optimum transmit signal using majorization theory. Interpretation of the results and simulation examples are presented in Section V. Concluding remarks are presented in Section VI.

#### Notation

For an integer  $Q$ ,  $\mathbf{I}_Q$  is a  $Q \times Q$  identity matrix. If  $\mathbf{X}$  is a  $Q \times K$  matrix, then its lowercase letter  $\mathbf{x} = \text{vec}(\mathbf{X})$  denotes the  $QK \times 1$  vector obtained by stacking columns of  $\mathbf{X}$ .  $\otimes$  denotes the Kronecker product [15].  $\mathbf{X}^*$ ,  $\mathbf{X}^T$ , and  $\mathbf{X}^H$  denote the complex conjugate, transpose, and Hermitian transpose of  $\mathbf{X}$ , respectively. The inverse, pseudo-inverse, trace, and determinant of  $\mathbf{X}$  are denoted by  $\mathbf{X}^{-1}$ ,  $\mathbf{X}^\dagger$ ,  $\text{tr}(\mathbf{X})$ , and  $\det(\mathbf{X})$ , respectively.  $\text{diag}([a_1, \dots, a_Q])$  is a  $Q \times Q$  diagonal matrix with elements  $a_1, \dots, a_Q$ .  $E(\cdot)$  denotes the expectation operator.  $(x)^+ = \max(0, x)$ .  $\mathbf{X}[m, n]$  denotes the entry in the  $m$ th row and  $n$ th column of  $\mathbf{X}$ .

## II. MIMO CHANNEL MODEL

Consider a narrowband frequency nonselective MIMO channel with  $P$  transmit and  $Q$  receive antennae. With  $k$  indicating discrete time, if  $\mathbf{s}(k)$  is the transmit vector of dimension  $P$ , then the  $Q$ -dimensional received signal  $\mathbf{x}(k)$  can be written as

$$\mathbf{x}(k) = \mathbf{H}(k)\mathbf{s}(k) + \mathbf{n}(k) \quad (1)$$

where  $\mathbf{H}(k)$  is the  $Q \times P$  channel matrix coupling the transmitter and receiver antennae.  $\mathbf{n}(k)$  is the  $Q$ -dimensional noise vector, which is assumed to be zero mean, complex white Gaussian with covariance matrix  $\sigma^2\mathbf{I}_Q$ . Assume the MIMO channel to be block fading, i.e.,  $\mathbf{H}(k) = \mathbf{H}$  for  $k = 1, \dots, T_c$ , and the channel is independent between different blocks of  $T_c$  symbols. Hence, we will suppress the index  $k$  in  $\mathbf{H}(k)$ . The channel gain between the  $n$ th receive and  $m$ th transmit antenna is the corresponding entry in the matrix  $\mathbf{H}$  denoted by  $\mathbf{H}[n, m]$ . The channel correlation is defined as  $\mathbf{R} = E(\mathbf{h}\mathbf{h}^H)$ , where  $\mathbf{h} = \text{vec}(\mathbf{H})$ . In this paper, we assume that the channel matrix can be expressed by the following canonical statistical model [23]:

$$\mathbf{H} = \mathbf{U}_R \mathbf{H}_V \mathbf{U}_T^H \quad (2)$$

where  $\mathbf{U}_T$  and  $\mathbf{U}_R$  are the transmit and receive unitary matrices, and the elements of  $\mathbf{H}_V$  are *uncorrelated* but not necessarily identically distributed.  $\mathbf{U}_T$  and  $\mathbf{U}_R$  correspond to the transmit and receive eigen-matrices (matrix of eigenvectors), respectively, i.e., the columns of  $\mathbf{U}_T$  and  $\mathbf{U}_R$  are the eigenvectors of  $E(\mathbf{H}^H\mathbf{H})$  and  $E(\mathbf{H}\mathbf{H}^H)$ , respectively. The correlation of the elements of  $\mathbf{H}_V$  is given by a *diagonal* matrix  $\mathbf{R}_V = E(\mathbf{h}_V\mathbf{h}_V^H)$ , where  $\mathbf{h}_V = \text{vec}(\mathbf{H}_V)$ . Note that we can write  $\mathbf{h} = \text{vec}(\mathbf{H}) = (\mathbf{U}_T^* \otimes \mathbf{U}_R)\mathbf{h}_V$ .  $\mathbf{R}$  and  $\mathbf{R}_V$  are related as

$$\mathbf{R} = (\mathbf{U}_T^* \otimes \mathbf{U}_R)\mathbf{R}_V(\mathbf{U}_T^* \otimes \mathbf{U}_R)^H. \quad (3)$$

In the following, we show that two important channel representations can be expressed in the above form.

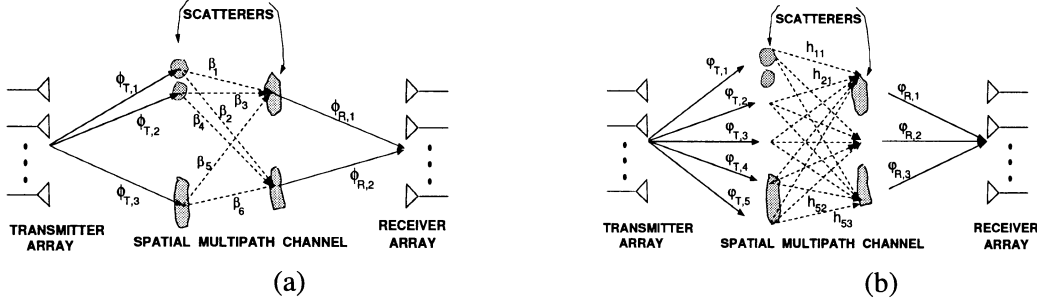


Fig. 1. (a). Physical channel modeling and (b). Virtual channel representation.

### A. ULAs and the Virtual Channel Representation

In [7], a *virtual* channel representation is proposed where the transmitter and receiver consist of ULAs of antennae. The virtual representation shows that for ULAs, the channel can be written in the form of (2), where  $\mathbf{U}_R$  and  $\mathbf{U}_T$  are discrete Fourier transform (DFT) matrices. In this section, we will elaborate on this a little more, but for details, see [7].

If  $d_T$  and  $d_R$  are the transmit and receive antenna spacings, then  $\mathbf{H}$  can be related to the physical propagation environment via the following array steering and response vectors:

$$\begin{aligned} \mathbf{a}_T(\theta_T) &= \frac{1}{\sqrt{P}} [1, \exp(-j2\pi\theta_T), \dots, \exp(-j2\pi(P-1)\theta_T)]^T \\ \mathbf{a}_R(\theta_R) &= \frac{1}{\sqrt{Q}} [1, \exp(-j2\pi\theta_R), \dots, \exp(-j2\pi(Q-1)\theta_R)]^T \end{aligned} \quad (4)$$

where  $\theta$  is the delay between the signals received at adjacent elements in the array due to a point source at the physical angle  $\phi$  measured relative to a horizontal axis; see Fig. 1(a). If  $\lambda$  is the wavelength of propagation, then  $\theta = (d/\lambda)\sin\phi$ . We will interpret  $\theta$  as a normalized angle. The vector  $\mathbf{a}_R(\theta_R)$  represents the signal response at the receiver array due to a point source in the direction  $\theta_R$ , whereas  $\mathbf{a}_T(\theta_T)$  represents array weights needed to transmit a beam focussed in the direction of  $\theta_T$ . The channel matrix for a widely used discrete path physical model, which is illustrated in Fig. 1(a), can be written as

$$\mathbf{H} = \sum_{l=1}^L \beta_l \mathbf{a}_R(\theta_{R,l}) \mathbf{a}_T^H(\theta_{T,l}) = \mathbf{A}_R \mathbf{H}_P \mathbf{A}_T^H. \quad (5)$$

In the above model, the transmitter and receiver are coupled via  $L$  propagation paths with  $\{\theta_{T,l}\}$  and  $\{\theta_{R,l}\}$  as the spatial angles seen by the transmitter and receiver, respectively, and  $\{\beta_l\}$  as the corresponding path gains. The matrices are defined as  $\mathbf{A}_R = [\mathbf{a}_R(\theta_{R,1}), \dots, \mathbf{a}_R(\theta_{R,L})]$  ( $Q \times L$ ),  $\mathbf{A}_T = [\mathbf{a}_T(\theta_{T,1}), \dots, \mathbf{a}_T(\theta_{T,L})]$  ( $P \times L$ ), and  $\mathbf{H}_P = \text{diag}(\beta_1, \dots, \beta_L)$ . Thus, this discrete model is linear in path gains but nonlinear in the spatial angles.

The linear virtual channel representation in [7] exploits the finite dimensionality of the spatial signal space due to a finite number of antennae and, hence, finite array apertures. Without loss of generality, assume  $P$  and  $Q$  to be odd and define  $\tilde{Q} = (Q-1)/2$  and  $\tilde{P} = (P-1)/2$ . The virtual channel representation is given by

$$\mathbf{H} = \sum_{q=-\tilde{Q}}^{\tilde{Q}} \sum_{p=-\tilde{P}}^{\tilde{P}} \mathbf{H}_V[q,p] \mathbf{a}_R(\tilde{\theta}_{R,q}) \mathbf{a}_T^H(\tilde{\theta}_{T,p}) = \tilde{\mathbf{A}}_R \mathbf{H}_V \tilde{\mathbf{A}}_T^H \quad (6)$$

where  $\tilde{\mathbf{A}}_R = [\mathbf{a}_R(\tilde{\theta}_{R,-\tilde{Q}}), \dots, \mathbf{a}_R(\tilde{\theta}_{R,\tilde{Q}})]$  ( $Q \times Q$ ) and  $\tilde{\mathbf{A}}_T = [\mathbf{a}_T(\tilde{\theta}_{T,-\tilde{P}}), \dots, \mathbf{a}_T(\tilde{\theta}_{T,\tilde{P}})]$  ( $P \times P$ ) are defined by the *fixed* virtual angles  $\tilde{\theta}_{R,q}$  and  $\tilde{\theta}_{T,p}$  and are full rank. The virtual channel representation is illustrated in Fig. 1(b). We assume that the spatial virtual angles are uniformly spaced [7]. As a result, the matrices  $\tilde{\mathbf{A}}_T$  and  $\tilde{\mathbf{A}}_R$  become DFT matrices and are, hence, unitary. Note that the virtual model is *linear* and is characterized by the virtual channel coefficients ( $\mathbf{H}_V$ ) since the spatial angles are fixed *a priori* (by the spatial resolution of the arrays).

Using (6), we can write

$$\mathbf{h} = (\tilde{\mathbf{A}}_T^* \otimes \tilde{\mathbf{A}}_R) \mathbf{h}_V \quad (7)$$

and

$$\mathbf{R} = (\tilde{\mathbf{A}}_T^* \otimes \tilde{\mathbf{A}}_R) \mathbf{R}_V (\tilde{\mathbf{A}}_T^* \otimes \tilde{\mathbf{A}}_R)^H. \quad (8)$$

An important property of the virtual representation is that the elements of  $\mathbf{H}_V$  are approximately *uncorrelated*, and hence,  $\mathbf{R}_V$  is approximately diagonal for any given channel correlation [7]. This shows that the virtual representation has the same form in (2) with  $\mathbf{U}_T = \tilde{\mathbf{A}}_T$  and  $\mathbf{U}_R = \tilde{\mathbf{A}}_R$ .

### B. Transmit and Receive Correlated Channel Model

In this channel model, it is assumed that the transmitter and receiver antenna arrays have correlated elements. The channel matrix can be written as

$$\mathbf{H} = \mathbf{\Sigma}_R^{1/2} \mathbf{H}_w \mathbf{\Sigma}_T^{1/2} \quad (9)$$

where the elements of  $\mathbf{H}_w$  are i.i.d. The matrices  $\mathbf{\Sigma}_T$  and  $\mathbf{\Sigma}_R$  are the transmit and receive array correlation matrices. The model was assumed in [1], [8], [9], and [16] and verified by measurements under certain environments in [10] and [17]. Special cases of the channel (only transmit correlation) are assumed in [6].

Denote the eigen value decompositions (EVD) of  $\mathbf{\Sigma}_T$  and  $\mathbf{\Sigma}_R$  as  $\mathbf{U}_T \mathbf{\Lambda}_T \mathbf{U}_T^H$  and  $\mathbf{U}_R \mathbf{\Lambda}_R \mathbf{U}_R^H$ , respectively. Using (9), we can write

$$\begin{aligned} \mathbf{H} &= \mathbf{U}_R \mathbf{\Lambda}_R^{1/2} \mathbf{U}_R^H \mathbf{H}_w \mathbf{U}_T \mathbf{\Lambda}_T^{1/2} \mathbf{U}_T^H \\ &= \mathbf{U}_R \mathbf{H}_V \mathbf{U}_T^H \end{aligned} \quad (10)$$

where the second equality arises from the following two observations. First, the elements of  $\mathbf{U}_R^H \mathbf{H}_w \mathbf{U}_T$  are still i.i.d., and second, the pre- and post-multiplication of diagonal matrices  $\mathbf{\Lambda}_R^{1/2}$  and  $\mathbf{\Lambda}_T^{1/2}$ , respectively, makes the elements of  $\mathbf{H}_V = \mathbf{\Lambda}_R^{1/2} \mathbf{U}_R^H \mathbf{H}_w \mathbf{U}_T \mathbf{\Lambda}_T^{1/2}$  uncorrelated with diagonal covariance matrix given by

$$\mathbf{R}_V = \mathbf{\Lambda}_T \otimes \mathbf{\Lambda}_R. \quad (11)$$

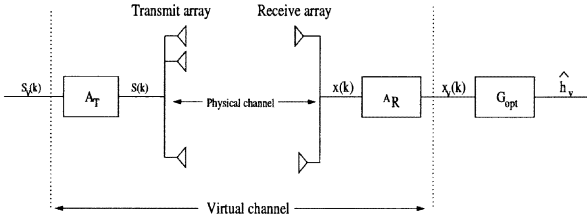


Fig. 2. Channel estimation in the virtual domain.

From (10), we can conclude that the transmit and receive correlated channel is a special case of our channel model with  $\mathbf{R}_V$  given in (11). It is interesting to note that when  $\Sigma_T$  and  $\Sigma_R$  are Toeplitz (which can be shown to be the case for ULAs [7]), then it can be shown that  $\mathbf{U}_T \approx \hat{\mathbf{A}}_T$  and  $\mathbf{U}_R \approx \hat{\mathbf{A}}_R$  since DFT matrices approximately diagonalize Toeplitz matrices. This shows that when  $\Sigma_T$  and  $\Sigma_R$  are Toeplitz, the model (9) is a special case of the virtual channel model (6). We note that the virtual representation provides the most general model for ULAs and includes (11) as a special case. Recent measurement results indicate that the product correlation in (11) may not be rich enough in general, and thus, an arbitrarily diagonal  $\mathbf{R}_V$  may be needed [18].

### C. Equivalent Channel Estimation

From (2), we can write

$$\mathbf{H}_V = \mathbf{U}_R^H \mathbf{H} \mathbf{U}_T \quad (12)$$

which indicates that  $\mathbf{H}$  and  $\mathbf{H}_V$  are unitarily equivalent. This implies that estimation of the MIMO channel  $\mathbf{H}$  can be equivalently performed by obtaining estimates of  $\mathbf{H}_V$ . From (1) and (2), we can write the received signal as

$$\mathbf{x}(k) = \mathbf{U}_R \mathbf{H}_V(k) \mathbf{U}_T^H \mathbf{s}(k) + \mathbf{n}(k). \quad (13)$$

In the eigen or virtual domain

$$\mathbf{x}_V(k) = \mathbf{H}_V(k) \mathbf{s}_V(k) + \mathbf{n}_V(k) \quad (14)$$

where  $\mathbf{x}_V = \mathbf{U}_R^H \mathbf{x}$  and  $\mathbf{s}_V = \mathbf{U}_T^H \mathbf{s}$  are the projections of the received and transmitted signals onto the receive and transmit eigen-matrices, respectively. Equation (14) provides an interesting interpretation of transmission in the virtual domain, which is illustrated in Fig. 2 for ULAs. Each element of  $\mathbf{x}_V$  corresponds to a signal received from the eigen-directions indicated by the receive eigen-matrix or the fixed virtual angle  $\tilde{\theta}_{R,q}$  in the case of ULAs. Similarly, each element of  $\mathbf{s}_V$  corresponds to a signal transmitted in the eigen-directions indicated by the transmit eigen-matrix or the fixed virtual angles  $\tilde{\theta}_{T,q}$  in the case of ULAs. The corresponding element in  $\mathbf{H}_V$  indicates the channel gain associated with the transmit and receive eigen-directions or virtual angles. Note that since  $\mathbf{U}_R$  is unitary,  $\mathbf{n}_V = \mathbf{U}_R \mathbf{n}$  is zero mean, white Gaussian with covariance  $\sigma^2 \mathbf{I}_Q$ . In the sequel, we will obtain physical insights from our analysis using the virtual representation in terms of the transmit and receive angles. However, our mathematical development applies to the general model in (2).

### D. Block Fading Channel Model

Let  $f_d$  be the two-sided Doppler spread of the MIMO channel and  $T$  be the symbol period. Then, the coherence block length is defined as  $T_c = \lfloor 1/f_d T \rfloor$  symbol periods. Hence, fast fading channels will have smaller  $T_c$  than slow fading channels. Assume the MIMO channel to be block fading, i.e.,  $\mathbf{H}(k) = \mathbf{H}$  for  $k = 1, \dots, T_c$ ,  $T_c > P$ , and the channel is independent between different blocks of  $T_c$  symbols. In addition, assume that the channel is estimated using a block of  $K$  symbols. In this paper, we are concerned with the design of the block training signal and estimating the optimal  $K$  for a given SNR and  $\mathbf{R}_V$ .

In general,  $N_t$  training blocks each of length  $K$  symbol periods can be used to improve channel estimates by averaging the  $N_t$  training blocks at the receiver. The channel is estimated from this averaged block, which exhibits an improved effective SNR (effective SNR = SNR  $\times N_t$ ). We can apply our framework at this effective SNR to determine the optimal  $K$  and the corresponding transmit block signal. If  $T_d$  data symbols are transmitted during the coherence time, then  $T_t + T_d = T_c$ , where  $T_t = N_t K$ . The tradeoff between  $T_d$  and  $T_t$  could be investigated along the lines of [19], in conjunction with this work.

Assuming that training symbols  $\mathbf{s}(k)$ ,  $k = 1, \dots, K$  are sent in a block mode, and denoting  $\mathbf{S} = [\mathbf{s}(1), \dots, \mathbf{s}(K)]$ , the block fading model is given by

$$\begin{aligned} \mathbf{X}_V &= [\mathbf{x}_V(1), \dots, \mathbf{x}_V(K)] \\ &= \mathbf{H}_V [\mathbf{s}_V(1), \dots, \mathbf{s}_V(K)] + [\mathbf{n}_V(1), \dots, \mathbf{n}_V(K)] \\ &= \mathbf{H}_V \mathbf{S}_V + \mathbf{N}_V. \end{aligned} \quad (15)$$

Stacking the columns of  $\mathbf{X}_V$ , we obtain

$$\tilde{\mathbf{x}}_V = \text{vec}(\mathbf{X}_V) = (\mathbf{S}_V^T \otimes \mathbf{I}_Q) \text{vec}(\mathbf{H}_V) + \text{vec}(\mathbf{N}_V) = \tilde{\mathbf{S}}_V \mathbf{h}_V + \tilde{\mathbf{n}}_V \quad (16)$$

where  $\tilde{\mathbf{x}}_V$  is a  $KQ \times 1$  vector, and we denote  $\tilde{\mathbf{S}}_V = (\mathbf{S}_V^T \otimes \mathbf{I}_Q) (KQ \times PQ)$ . Using (16), we proceed with the estimation of  $\mathbf{h}_V$ , which is a  $PQ$  vector. Clearly, since the maximum number of unknowns<sup>3</sup> in  $\mathbf{h}_V$  is  $PQ$ , we need to transmit a block of  $K \leq P$  symbols [19]. In the following discussion, we start by letting  $K$  take its maximum value  $K = P$ , and we will see that the optimal transmit signal design exactly identifies the true block length for correlated channels by assigning zero power to some of the  $P$  transmit virtual angles.

## III. MMSE AND MAP ESTIMATION

The model (16) is linear in  $\mathbf{h}_V$  and Gaussian. Hence, it can be shown that the linear MMSE estimate, the MMSE estimate, and the maximum *a posteriori* (MAP) estimate are identical [5]. In this paper, we assume that the covariance matrix  $\mathbf{R}_V = E(\mathbf{h}_V \mathbf{h}_V^H)$  (or equivalently  $\mathbf{R} = E(\mathbf{h} \mathbf{h}^H)$ ) is known. We assume that  $\mathbf{R}_V$  is full rank; however, this condition will be relaxed later (see remarks at the end of Section IV-A).

The linear MMSE estimator minimizes the error given by

$$\text{MSE} = E[\|\mathbf{h}_V - \hat{\mathbf{h}}_V\|^2]. \quad (17)$$

<sup>3</sup>The number of unknowns in  $\mathbf{h}_V$  may be smaller in correlated channels. If the prior variance of a given element of  $\mathbf{h}_V$  is zero, then it implies that the element is itself zero.

Using the orthogonality principle, the resulting linear estimate is

$$\hat{\mathbf{h}}_V = \mathbf{G}_{\text{opt}} \tilde{\mathbf{x}}_V \quad (18)$$

where  $\mathbf{G}_{\text{opt}}$  is a  $PQ \times PQ$  matrix given by

$$\begin{aligned} \mathbf{G}_{\text{opt}} &= \arg \min_{\mathbf{G}} E[\|\mathbf{h}_V - \mathbf{G}\tilde{\mathbf{x}}_V\|^2] \\ &= \mathbf{R}_V \tilde{\mathbf{S}}_V^H (\tilde{\mathbf{S}}_V \mathbf{R}_V \tilde{\mathbf{S}}_V^H + \sigma^2 \mathbf{I})^{-1}. \end{aligned} \quad (19)$$

The error covariance matrix is given by

$$\begin{aligned} \mathbf{C}_e &= E[(\mathbf{h}_V - \hat{\mathbf{h}}_V)(\mathbf{h}_V - \hat{\mathbf{h}}_V)^H] \\ &= \mathbf{R}_V - \mathbf{R}_V \tilde{\mathbf{S}}_V^H (\tilde{\mathbf{S}}_V \mathbf{R}_V \tilde{\mathbf{S}}_V^H + \sigma^2 \mathbf{I}_{QK})^{-1} \tilde{\mathbf{S}}_V \mathbf{R}_V \\ &= (\mathbf{R}_V^\dagger + \frac{1}{\sigma^2} \tilde{\mathbf{S}}_V^H \tilde{\mathbf{S}}_V)^{-1}. \end{aligned} \quad (20)$$

The minimum MSE is

$$\text{MMSE} = \text{tr}(\mathbf{C}_e(\tilde{\mathbf{S}}_V)). \quad (21)$$

In fact, the posterior distribution of  $\mathbf{h}_V$  is Gaussian [5], i.e.,

$$p(\mathbf{h}_V | \mathbf{x}_V, \tilde{\mathbf{S}}_V) = \mathcal{N}(\mathbf{G}_{\text{opt}} \mathbf{x}_V, \mathbf{C}_e). \quad (22)$$

The conditional mutual information (CMI) between the received signal and the channel  $\mathbf{h}_V$  is given by

$$\begin{aligned} \text{CMI}(\tilde{\mathbf{S}}_V) &= \mathcal{I}(\mathbf{h}_V; \mathbf{x}_V | \tilde{\mathbf{S}}_V) = \mathcal{H}(\mathbf{h}_V) - \mathcal{H}(\mathbf{h}_V | \mathbf{x}_V, \tilde{\mathbf{S}}_V) \\ &= \log \det(\mathbf{R}_V) - \log \det(\mathbf{C}_e) \\ &= \log \det(\mathbf{I} + \frac{1}{\sigma^2} \tilde{\mathbf{S}}_V \mathbf{R}_V \tilde{\mathbf{S}}_V^H) \end{aligned} \quad (23)$$

where  $\mathcal{I}(x; y)$  denotes the mutual information between  $x$  and  $y$ , and  $\mathcal{H}(x)$  denotes the entropy of  $x$  [12].

#### IV. OPTIMUM SIGNAL DESIGN

In this section, we consider the design of the optimum transmit block signal  $\tilde{\mathbf{S}}_V$  (or equivalently  $\mathbf{S}_V$ ) with respect to two criteria: minimization of the MMSE (21) and maximization of the mutual information (23) between the channel and received signal conditioned on the transmitted block signal. We state the two optimization problems as follows:

$$\begin{aligned} \text{OP}_{\text{MMSE}} : \min_{\tilde{\mathbf{S}}_V} \quad & \text{tr} \left( \mathbf{R}_V^\dagger + \frac{1}{\sigma^2} \tilde{\mathbf{S}}_V^H \tilde{\mathbf{S}}_V \right)^{-1} \\ \text{s.t.} \quad & \text{tr} \left( \tilde{\mathbf{S}}_V^H \tilde{\mathbf{S}}_V \right) \leq P\beta \end{aligned} \quad (24)$$

and

$$\begin{aligned} \text{OP}_{\text{CMI}} : \max_{\tilde{\mathbf{S}}_V} \quad & \log \det \left( \mathbf{I} + \frac{1}{\sigma^2} \tilde{\mathbf{S}}_V \mathbf{R}_V \tilde{\mathbf{S}}_V^H \right) \\ \text{s.t.} \quad & \text{tr} \left( \tilde{\mathbf{S}}_V^H \tilde{\mathbf{S}}_V \right) \leq P\beta \end{aligned} \quad (25)$$

where  $\beta$  is the total transmitted power, and  $P$  is the number of transmit antennae. Note that the constraint  $\text{tr}(\tilde{\mathbf{S}}_V^H \tilde{\mathbf{S}}_V) \leq P\beta$  is equivalent to the finite power constraint  $\text{tr}(\mathbf{S}_V^H \mathbf{S}_V) = \text{tr}(\mathbf{S}^H \mathbf{S}) \leq \beta$  since  $\tilde{\mathbf{S}}_V = (\mathbf{S}_V^T \otimes \mathbf{I}_Q)$ .

The MMSE estimate is optimal in the sense that it is also the MAP estimate since the channel estimation problem here is linear and Gaussian. One motivation for minimizing the MMSE arises from [19], where optimal training is considered for estimation of MIMO channels, and it is shown that minimizing the MMSE is optimal from the point of view of maximizing

capacity. Second, the impact of channel estimation on overall bit error rate performance can be assessed, to a first approximation, by including the error variance (MMSE) in the background noise variance; see, for example, [20].

While the MMSE criterion minimizes the trace of  $\mathbf{C}_e$ , the CMI criterion is motivated by the minimization of the determinant of  $\mathbf{C}_e$ . From (23), we have  $\text{CMI}(\tilde{\mathbf{S}}_V) = \log \det(\mathbf{R}_V) - \log \det(\mathbf{C}_e)$ . Thus, while the MMSE criterion minimizes the sum of the eigenvalues (trace) of  $\mathbf{C}_e$ , the CMI criterion minimizes the sum of the log of the eigenvalues (logarithm of determinant) of  $\mathbf{C}_e$ . The CMI criterion is motivated from an information-theoretic argument and suggests the maximization of the mutual information between the unknown channel and received signal as a function of transmitted training symbols. A dual of this criterion is where the channel is assumed known at the receiver while symbols are unknown. In this usual context, the mutual information between the transmitted symbols and received signal is considered from a coherent capacity perspective. Maximizing CMI is equivalent to minimizing the second term on the right-hand side in (23), i.e., the entropy or *uncertainty* in  $\mathbf{h}_V$  given  $\mathbf{x}_V$  by proper design of  $\tilde{\mathbf{S}}_V$ . In [11], these two criteria are considered for the MISO case.

Note that the optimal problems posed in the virtual domain are equivalent to formulating them in the antenna domain. This is evident by noting that since  $\mathbf{H}$  and  $\mathbf{H}_V$  are unitarily equivalent, the MMSEs are equal, i.e.,  $\text{MSE} = E[\|\mathbf{h}_V - \hat{\mathbf{h}}_V\|^2] = E[\|\mathbf{h} - \hat{\mathbf{h}}\|^2]$ , where  $\hat{\mathbf{h}}$  is the equivalent linear estimate in the antenna domain. In addition, denoting  $\tilde{\mathbf{S}} = \mathbf{S}^T \otimes \mathbf{I}_Q$ , one can write  $\text{CMI} = \log \det(\mathbf{I} + (1/\sigma^2) \tilde{\mathbf{S}}_V \mathbf{R}_V \tilde{\mathbf{S}}_V^H) = \log \det(\mathbf{I} + (1/\sigma^2) \tilde{\mathbf{S}} \mathbf{R} \tilde{\mathbf{S}}^H)$  implying equivalence for the CMI condition.

##### A. Structure of Optimal $\tilde{\mathbf{S}}_V$

We develop the transmit signal design using the singular value decomposition (SVD) of the transmitted block matrix. Denote the SVDs of  $\mathbf{S}_V^T = \mathbf{U}_S \Lambda_S \mathbf{V}_S^H$  ( $P \times P$ ) and  $\tilde{\mathbf{S}}_V = \mathbf{U}_{\tilde{S}} \Lambda_{\tilde{S}} \mathbf{V}_{\tilde{S}}^H$  ( $PQ \times PQ$ ), where  $\mathbf{U}_S$ ,  $\mathbf{V}_S$ ,  $\mathbf{U}_{\tilde{S}}$ , and  $\mathbf{V}_{\tilde{S}}$  are unitary matrices, and  $\Lambda_S$  and  $\Lambda_{\tilde{S}}$  are diagonal matrices. Since  $\tilde{\mathbf{S}}_V = (\mathbf{S}_V^T \otimes \mathbf{I}_Q)$ , it follows that  $\mathbf{U}_{\tilde{S}} = \mathbf{U}_S \otimes \mathbf{I}_Q$ ,  $\Lambda_{\tilde{S}} = \Lambda_S \otimes \mathbf{I}_Q$ , and  $\mathbf{V}_{\tilde{S}} = \mathbf{V}_S \otimes \mathbf{I}_Q$ .

*Lemma 1:* The transmit block signal that optimizes  $\text{OP}_{\text{MMSE}}$  and  $\text{OP}_{\text{CMI}}$  has a structure given by  $\tilde{\mathbf{S}}_V = \Lambda_{\tilde{S}} \mathbf{V}_{\tilde{S}}^H$ .

*Proof:* Both the optimization criteria are a function of  $\tilde{\mathbf{S}}_V$  through the error covariance matrix [see (21) and (23)], which can be written as

$$\mathbf{C}_e = \left( \mathbf{R}_V^\dagger + \frac{1}{\sigma^2} \tilde{\mathbf{S}}_V^H \tilde{\mathbf{S}}_V \right)^{-1} = \left( \mathbf{R}_V^\dagger + \frac{1}{\sigma^2} \mathbf{V}_{\tilde{S}} \Lambda_{\tilde{S}}^H \Lambda_{\tilde{S}} \mathbf{V}_{\tilde{S}}^H \right)^{-1} \quad (26)$$

where we have used the SVD of  $\tilde{\mathbf{S}}_V$ . This indicates that  $\mathbf{C}_e$  is independent of  $\mathbf{U}_{\tilde{S}}$ , and hence, we can choose  $\mathbf{U}_{\tilde{S}} = \mathbf{I}_{PQ}$  without loss of generality, which proves the above claim. ■

Assuming this structure, we show in the following development that the optimal transmit signal *diagonalizes* the covariance of the channel. To prove this result, we use concepts from the theory of majorization and Schur-convex functions, which are discussed in the Appendix.

1) *Minimizing the MMSE:* We design the transmit signal to minimize the MMSE by solving  $\text{OP}_{\text{MMSE}}$  (24), which is only a

function of  $\mathbf{V}_{\tilde{\mathcal{S}}}$  and  $\Lambda_{\tilde{\mathcal{S}}}$  due to Lemma 1. We have the following theorem.

*Theorem 1:* Consider the constrained optimization problem  $\text{OP}_{\text{MMSE}}$  in (24). It follows that the globally optimal solution has a diagonal structure, that is

$$\tilde{\mathbf{S}}_{V,\text{opt}} = \tilde{\Lambda}_M \quad (27)$$

where  $\tilde{\Lambda}_M \in \mathbb{R}^{PQ \times PQ}$ , and the subscript  $M$  denotes ‘‘MMSE optimal.’’

*Proof:* Using Lemma 1, we can write

$$\begin{aligned} \text{tr}(\mathbf{C}_e) &= \text{tr} \left( \left( \mathbf{R}_V^\dagger + \frac{1}{\sigma^2} \mathbf{V}_{\tilde{\mathcal{S}}} \Lambda_{\tilde{\mathcal{S}}}^H \Lambda_{\tilde{\mathcal{S}}} \mathbf{V}_{\tilde{\mathcal{S}}}^H \right)^{-1} \right) \\ &= \text{tr} \left( \left( \mathbf{V}_{\tilde{\mathcal{S}}}^H \mathbf{R}_V^\dagger \mathbf{V}_{\tilde{\mathcal{S}}} + \frac{1}{\sigma^2} \Lambda_{\tilde{\mathcal{S}}}^H \Lambda_{\tilde{\mathcal{S}}} \right)^{-1} \right). \end{aligned}$$

Denote  $\mathbf{A} := \mathbf{V}_{\tilde{\mathcal{S}}}^H \mathbf{R}_V^\dagger \mathbf{V}_{\tilde{\mathcal{S}}} + (1/\sigma^2) \Lambda_{\tilde{\mathcal{S}}}^H \Lambda_{\tilde{\mathcal{S}}}$  and the eigenvalues of  $\mathbf{A}$  as  $\lambda_i(\mathbf{A})$ , and let  $\boldsymbol{\lambda}(\mathbf{A}) = (\lambda_1(\mathbf{A}), \dots, \lambda_{PQ}(\mathbf{A}))$ . Then, the MMSE can be written as

$$f(\boldsymbol{\lambda}(\mathbf{A})) := \text{tr}(\mathbf{A}^{-1}) = \sum_{i=1}^{PQ} \frac{1}{\lambda_i(\mathbf{A})}. \quad (28)$$

Since  $\phi(\lambda_i(\mathbf{A})) := 1/\lambda_i(\mathbf{A})$  s.t.  $\lambda_i(\mathbf{A}) \in \mathbb{R}^+$  is a convex function, by virtue of Theorem 4 in the Appendix,  $f(\boldsymbol{\lambda}(\mathbf{A}))$  is Schur-convex. We invoke Theorem 3 in the Appendix to conclude that

$$f(\boldsymbol{\lambda}(\mathbf{A})) \geq f(\mathbf{d}(\mathbf{A})) \quad (29)$$

where  $\mathbf{d}(\mathbf{A})$  is the vector of diagonal elements of  $\mathbf{A}$ . The lower bound in the above inequality is achieved when  $\boldsymbol{\lambda}(\mathbf{A}) = \mathbf{d}(\mathbf{A})$  or  $\mathbf{A}$  is a diagonal matrix. Clearly, this is true when  $\mathbf{V}_{\tilde{\mathcal{S}}}^H \mathbf{R}_V^\dagger \mathbf{V}_{\tilde{\mathcal{S}}}$  is diagonal. Since  $\mathbf{V}_{\tilde{\mathcal{S}}}$  is a unitary matrix, it must be equal to the eigen-matrix (matrix of eigenvectors) of  $\mathbf{R}_V^\dagger$  or a permutation thereof. Now, since  $\mathbf{R}_V$  is a diagonal matrix, its eigen-matrix is  $\mathbf{I}_{PQ}$ . Thus, we can write  $\mathbf{V}_{\tilde{\mathcal{S}}} = \mathbf{P}$ , where  $\mathbf{P}$  is a permutation matrix, and

$$\begin{aligned} \tilde{\mathbf{S}}_{V,\text{opt}} = \tilde{\Lambda}_M &= \arg \min_{\Lambda_{\tilde{\mathcal{S}}}} \text{tr} \left( \mathbf{R}_V^\dagger + \frac{1}{\sigma^2} \Lambda_{\tilde{\mathcal{S}}}^H \Lambda_{\tilde{\mathcal{S}}} \right)^{-1} \\ \text{s.t.} & \text{tr}(\Lambda_{\tilde{\mathcal{S}}}^H \Lambda_{\tilde{\mathcal{S}}}) \leq P\beta. \end{aligned} \quad (30)$$

Without loss of generality, we can choose  $\mathbf{V}_{\tilde{\mathcal{S}}} = \mathbf{I}_{PQ}$  since a permutation in  $\mathbf{V}_{\tilde{\mathcal{S}}}$  would only change the ordering of  $\tilde{\Lambda}_M$ . This concludes the proof. ■

2) *Maximizing the CMI:* We now show that the optimal transmit signal that maximizes CMI ( $\text{OP}_{\text{CMI}}$  in (25)) also has a diagonal structure.

*Theorem 2:* Considering the constrained optimization problem  $\text{OP}_{\text{CMI}}$  in (25), it follows that the globally optimal solution has a diagonal structure. That is

$$\tilde{\mathbf{S}}_{V,\text{opt}} = \tilde{\Lambda}_C \quad (31)$$

where  $\tilde{\Lambda}_C \in \mathbb{R}^{PQ \times PQ}$ , and the subscript  $C$  denotes ‘‘CMI optimal.’’

*Proof:* Let  $\mathbf{B} := \mathbf{I} + (1/\sigma^2) \tilde{\mathbf{S}}_V \mathbf{R}_V \tilde{\mathbf{S}}_V^H$ , and denote the eigenvalues and main diagonal elements of  $\mathbf{B}$  as  $\lambda_i(\mathbf{B})$  and  $d_i(\mathbf{B})$ . In addition, let  $\boldsymbol{\lambda}(\mathbf{B}) = (\lambda_1(\mathbf{B}), \dots, \lambda_{PQ}(\mathbf{B}))$  and

$\mathbf{d}(\mathbf{B}) = (d_1(\mathbf{B}), \dots, d_{PQ}(\mathbf{B}))$ . Then, the CMI can be written as

$$g(\boldsymbol{\lambda}(\mathbf{B})) := \log \det(\mathbf{B}) = \sum_{i=1}^{PQ} \log(\lambda_i(\mathbf{B})). \quad (32)$$

Since  $\phi(d_i) := \log(d_i)$  is concave,  $g(\boldsymbol{\lambda}(\mathbf{B}))$  is Schur-concave due to Corollary 1 in the Appendix. As a result, from Theorem 3, we have

$$g(\boldsymbol{\lambda}(\mathbf{B})) \leq g(\mathbf{d}(\mathbf{B})) \quad (33)$$

and the upper bound is achieved when  $\boldsymbol{\lambda}(\mathbf{B}) = \mathbf{d}(\mathbf{B})$ . This is true when  $\mathbf{B}$  is diagonal or  $\tilde{\mathbf{S}}_V \mathbf{R}_V \tilde{\mathbf{S}}_V^H = \Lambda_{\tilde{\mathcal{S}}} \mathbf{V}_{\tilde{\mathcal{S}}}^H \mathbf{R}_V \mathbf{V}_{\tilde{\mathcal{S}}} \Lambda_{\tilde{\mathcal{S}}}^H$  is diagonal, where we have used Lemma 1. This implies that  $\mathbf{V}_{\tilde{\mathcal{S}}}^H \mathbf{R}_V \mathbf{V}_{\tilde{\mathcal{S}}}$  is diagonal, and using arguments similar to those in proof of Theorem 1, we have  $\mathbf{V}_{\tilde{\mathcal{S}}} = \mathbf{I}_{PQ}$  and

$$\begin{aligned} \tilde{\mathbf{S}}_{V,\text{opt}} = \tilde{\Lambda}_C &= \arg \max_{\Lambda_{\tilde{\mathcal{S}}}} \log \det(\mathbf{I} + \Lambda_{\tilde{\mathcal{S}}} \mathbf{R}_V \Lambda_{\tilde{\mathcal{S}}}^H) \\ \text{s.t.} & \text{tr}(\Lambda_{\tilde{\mathcal{S}}}^H \Lambda_{\tilde{\mathcal{S}}}) \leq P\beta. \end{aligned} \quad (34)$$

This completes the proof. ■

*Remark:* Care must be taken when  $\mathbf{R}_V$  is rank deficient. When  $\mathbf{R}_V$  is rank deficient, let  $\tilde{\mathbf{R}}_V = \mathbf{R}_V + \epsilon \mathbf{I}$  such that  $\epsilon$  is small so that  $\tilde{\mathbf{R}}_V$  is not rank deficient. Then, the results in (20), (23), (30), and (34) hold true for  $\tilde{\mathbf{R}}_V$  substituted instead of  $\mathbf{R}_V$ . Once (30) and (34) are obtained with this substitution, we can let  $\epsilon \rightarrow 0$  due to the continuity of the functions involved.

## B. Optimal Estimator Structure

Recall from Section IV-A that the SVDs of  $\mathbf{S}_V^T = \mathbf{U}_S \Lambda_S \mathbf{V}_S^H$  and  $\tilde{\mathbf{S}}_V = \mathbf{U}_{\tilde{\mathcal{S}}} \Lambda_{\tilde{\mathcal{S}}} \mathbf{V}_{\tilde{\mathcal{S}}}^H$ , together with  $\tilde{\mathbf{S}}_V = (\mathbf{S}_V^T \otimes \mathbf{I}_Q)$  imply that  $\mathbf{U}_{\tilde{\mathcal{S}}} = \mathbf{U}_S \otimes \mathbf{I}_Q$ ,  $\Lambda_{\tilde{\mathcal{S}}} = \Lambda_S \otimes \mathbf{I}_Q$  and  $\mathbf{V}_{\tilde{\mathcal{S}}} = \mathbf{V}_S \otimes \mathbf{I}_Q$ . From Lemma 1 and Theorems 1 and 2, the optimal transmit signal has  $\mathbf{U}_{\tilde{\mathcal{S}}} = \mathbf{I}_{PQ}$  and  $\mathbf{V}_{\tilde{\mathcal{S}}} = \mathbf{I}_{PQ}$ , and  $\Lambda_{\tilde{\mathcal{S}}}$  satisfies (30) and (34). This implies that  $\mathbf{U}_S = \mathbf{I}_P$  and  $\mathbf{V}_S = \mathbf{I}_P$  are optimal.

Now, the optimal transmit block signal  $\tilde{\mathbf{S}}_{V,\text{opt}} = \tilde{\Lambda}_M (\tilde{\Lambda}_C)$  according to the MMSE (CMI) criterion. This implies that the optimal  $\tilde{\mathbf{S}}_{V,\text{opt}} = \Lambda_M (\Lambda_C)$  according to MMSE (CMI) criterion, where  $\tilde{\Lambda}_M = \Lambda_M \otimes \mathbf{I}_Q$ , and  $\tilde{\Lambda}_C = \Lambda_C \otimes \mathbf{I}_Q$ . We will denote the power transmitted along the  $i$ th virtual angle as  $\beta_i = |\Lambda_S(i, i)|^2$ ,  $i = 1, \dots, P$ . Using (30) and (34), we have

$$\begin{aligned} \Lambda_M &= \arg \min_{\Lambda_S} \text{MMSE}(\Lambda_S) \\ &= \arg \min_{\Lambda_S} \sum_{i=1}^P \sum_{j=1}^Q \left( \frac{\sigma^2 \mathbf{R}_V[(i-1)Q+j, (i-1)Q+j]}{\sigma^2 + \mathbf{R}_V[(i-1)Q+j, (i-1)Q+j] \beta_i} \right) \end{aligned} \quad (35)$$

and

$$\begin{aligned} \Lambda_C &= \arg \max_{\Lambda_S} \text{CMI}(\Lambda_S) \\ &= \arg \max_{\Lambda_S} \sum_{i=1}^P \sum_{j=1}^Q \log \left( 1 + \frac{\mathbf{R}_V[(i-1)Q+j, (i-1)Q+j] \beta_i}{\sigma^2} \right) \end{aligned} \quad (36)$$

subject to the constraint  $\sum_{i=1}^P \beta_i \leq \beta$ .

We can now conclude that the optimal transmit signal is a block diagonal signal (in the virtual domain). The optimal signal

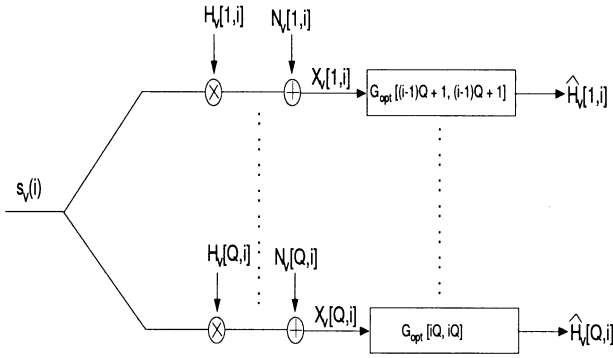


Fig. 3. "Diagonalizing" interpretation (in the virtual domain) of the optimal estimator. The  $i$ th transmit beam is used to estimate the  $i$ th column of  $\mathbf{H}_V$ . Further, each element of the  $i$ th column is estimated independently.

structure specifies that during the  $P$  block transmission, at each time instant  $i \in 1, \dots, P$ , the signal is transmitted along the  $i$ th virtual transmit angle, with the powers specified by  $\beta_i^{(M)} = |\Lambda_M[i, i]|^2$  ( $\beta_i^{(C)} = |\Lambda_C[i, i]|^2$ ) according to the MMSE (CMI) criterion.

We can make some more interesting observations. Due to the diagonal structure of  $\mathbf{S}_V$ , the linear receive processor  $\mathbf{G}_{\text{opt}}$  (19) and the error covariance matrix  $\mathbf{C}_e$  (20) become diagonal, which enables *independent* processing at the receiver. This is illustrated in Fig. 3. Let  $\Lambda_{\text{opt}}$  denote  $\Lambda_M$  or  $\Lambda_C$  depending on the criterion. Then the channel estimate can be written as

$$\begin{aligned} & \hat{\mathbf{h}}_V[(i-1)Q+j] \\ &= \left( \frac{\mathbf{R}_V[(i-1)Q+j, (i-1)Q+j] \Lambda_{\text{opt}}^H[i, i]}{\sigma^2 + \mathbf{R}_V[(i-1)Q+j, (i-1)Q+j] |\Lambda_{\text{opt}}[i, i]|^2} \right) \\ & \quad \cdot \mathbf{x}_V[(i-1)Q+j], \\ & j = 1, \dots, Q; i = 1, \dots, P. \end{aligned} \quad (37)$$

From this equation, note that the  $i$ th transmission allows us to estimate the  $Q$  elements in the  $i$ th column of  $\mathbf{H}_V$ , i.e.,  $(\mathbf{h}_V[(i-1)Q+1], \dots, \mathbf{h}_V[iQ])$ . During the block transmission, the scattering environment is scanned sequentially to estimate each column of  $\mathbf{H}_V$ . This *scanning* of the virtual domain is illustrated in Fig. 4. The individual error variance is given by

$$\begin{aligned} & E(|\mathbf{h}_V[(i-1)Q+j] - \hat{\mathbf{h}}_V[(i-1)Q+j]|^2) \\ &= \mathbf{C}_e[(i-1)Q+j, (i-1)Q+j] \\ &= \frac{\sigma^2 \mathbf{R}_V[(i-1)Q+j, (i-1)Q+j]}{\sigma^2 + \mathbf{R}_V[(i-1)Q+j, (i-1)Q+j] |\Lambda_{\text{opt}}(i, i)|^2} \\ & j = 1, \dots, Q; i = 1, \dots, P. \end{aligned} \quad (38)$$

*Remark:* The MMSE and CMI optimization metrics minimize the sum and product of the error covariances of the individual channel coefficients. We note that the theory of majorization can also be applied to ensure that all coefficients are estimated with the same accuracy. However, the above results show that not all channel coefficients are equally important. In essence, those channel coefficients whose power is small compared with the background noise are not important from a communication viewpoint, and the transmit power is better utilized in channel coefficients that exhibit a higher SNR.

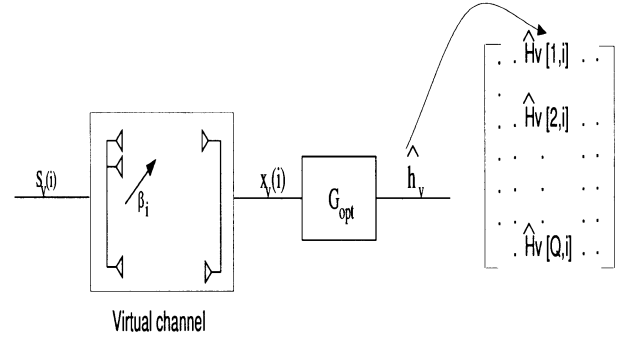


Fig. 4. Physical interpretation of channel estimation. "Scanning" the scattering environment in the virtual domain. The  $i$ th beam is used to estimate the  $i$ th column of  $\mathbf{H}_V$ .

### C. Water Filling Solution

The constrained nonlinear optimizations in (35) and (36) are the so called "water filling" problems and can be solved using Lagrange multipliers and using the Kuhn-Tucker conditions to verify that the solutions are non-negative. However, for the general case of  $P$  transmit and  $Q$  receive antennae, we have not been able to find a closed-form solution, and hence, it has to be obtained numerically. In the following, we obtain approximate closed-form solutions in the low SNR and high SNR regions to obtain some insight. Subsequently, some special cases are also considered, where we obtain closed-form solutions.

In the following discussion, the channel coefficient  $\mathbf{H}_V[j, i]$  is defined as *active* if  $E(|\mathbf{H}_V[j, i]|^2) > \epsilon$  for some prescribed  $\epsilon > 0$ . Thus, the set of active elements can be determined *a priori* by examining  $\mathbf{R}_V$ . In addition, a column of the channel matrix  $\mathbf{H}_V$  is defined as an *active column* if it contains at least one active element. Let  $A \subset \{1, \dots, P\}$  be the set of active columns. Then, for any  $i \in A$ ,  $Q_i$  is the number of active elements in the  $i$ th column. We define the transmit signal-to-noise ratio (TSNR) as the ratio of the transmitted signal power to the noise power  $\beta/\sigma^2$  and the *per virtual angle pair* received signal-to-noise ratio (RSNR) as the ratio of the received signal power to the noise power  $\text{RSNR}(i, j) = (\mathbf{R}_V[(i-1)Q+j, (i-1)Q+j] \beta_i)/\sigma^2$  between the  $i$ th transmit and  $j$ th receive angle pair for  $i = 1, \dots, P; j = 1, \dots, Q$ .

1) *Limiting Solution in High RSNR Regime:* Consider the high RSNR case, where  $\text{RSNR}(i, j) \gg 1$  for the set of active channel coefficients. Using Lagrange multipliers, it is shown in the Appendix that for high RSNR case, MMSE and CMI criteria assign power according to

$$\text{MMSE: } \beta_i = \frac{\sqrt{Q_i} \beta}{\sum_{l=1}^P \sqrt{Q_l}} \quad i = 1, \dots, P, \quad (39)$$

$$\text{and CMI: } \beta_i = \frac{Q_i \beta}{\sum_{l=1}^P Q_l} \quad i = 1, \dots, P \quad (40)$$

respectively. Thus, the CMI (MMSE) criterion assigns power to the transmit beams in proportion to the sum (square root of the sum) of the active elements they couple with at the receiver.

*Remark:* In the extreme case, when all elements of  $\mathbf{H}_V$  are active, then *equal* power is distributed to all transmit beams for both criteria.

2) *Limiting Solution in Low RSNR Regime:* Consider the low RSNR case, where  $\text{RSNR}(i, j) \ll 1, \forall i, j$ . Using Lagrange multipliers, it is shown in the Appendix that the MMSE and CMI criteria assign all the power  $\beta$  to the  $k$ th transmit beam such that

$$\text{MMSE} : k = \arg \max_i \sum_{j=1}^Q \mathbf{R}_V^2[(i-1)Q+j, (i-1)Q+j] \quad (41)$$

$$\text{and CMI} : k = \arg \max_i \sum_{j=1}^Q \mathbf{R}_V[(i-1)Q+j, (i-1)Q+j] \quad (42)$$

respectively. Thus, at low RSNR, the CMI (MMSE) criterion assigns all the power to that virtual transmit angle for which the sum (sum of squares) of the variances of the corresponding virtual receive elements is maximum.

From the extreme cases, we conclude that the number of transmit beams to be sent and, hence, the block length  $K$  depends on the SNR. For high SNR,  $K$  is equal to the number of active columns, whereas for low SNR,  $K = 1$ , and all the power is transmitted in the strongest transmit direction. For medium SNR,  $1 \leq K \leq P$ , and the powers will be determined by the water filling criteria. In addition, note that for i.i.d. channels, equal power will be assigned to all transmit beams, irrespective of the SNR.

3) *MISO ( $Q = 1$ ) Case:* In this case, it can be shown using Lagrange multipliers that the power assignment is identical for both criteria and is given by

$$\beta_i = \left( \frac{\beta}{P} + \frac{1}{P} \sum_{j=1}^P \frac{\sigma^2}{\mathbf{R}_V[j, j]} - \frac{\sigma^2}{\mathbf{R}_V[i, i]} \right)^+ \quad (43)$$

It is evident from the above equation that for high RSNR, the asymptotic equipower assignment in (39) and (40) is optimal. Note that the sum of the second and third terms in (43) is negative for a particular transmit angle if the inverse of its RSNR is greater than the average of the inverses of all the RSNRs. This is illustrated in Fig. 5. Hence, as the total power  $\beta$  is reduced, the weakest angle will get no power, and the power will be assigned among the remaining angles. Eventually, as  $\beta$  decreases (which will decrease the RSNR), power will be assigned to the highest RSNR angle, as indicated in (41) and (42).

We note that in [21], an optimal transmit diversity scheme is considered for a multiple transmit and single receive antenna system, where each transmit antenna is allocated a distinct CDMA code. The transmitted power is assigned to different antennae to minimize the overall bit error rate for a BPSK modulation scheme. It is observed that the optimal allocation obtained is identical to (43). In [11], optimal transmit diversity precoder design is discussed for a similar system. At the receiver, maximal ratio combining using the estimated channel is used to maximize the SNR. It is suggested that accurate channel estimates will make the MRC scheme effective, and hence,

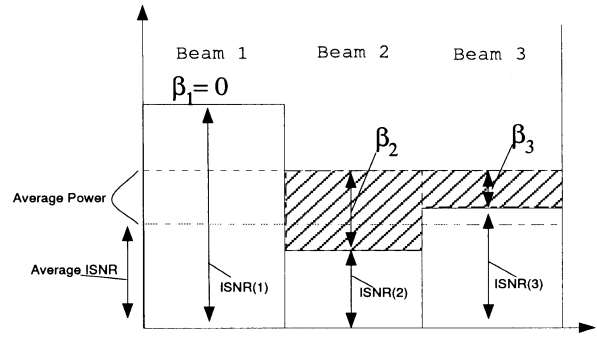


Fig. 5. Illustration of power allocation in (43).

power assignment is considered from a minimum MMSE criterion. Again, the optimal power assignment is identical to (43). We should note, however, that the aim in this work is optimal channel estimation, and our design is not motivated by transmit diversity considerations.

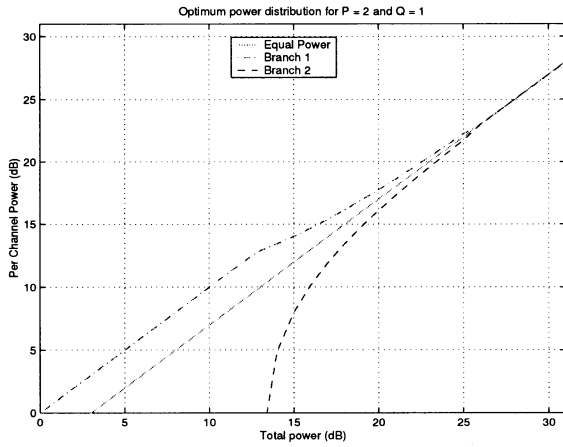
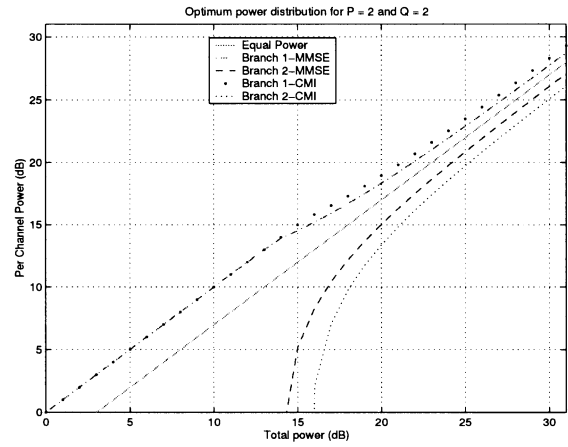
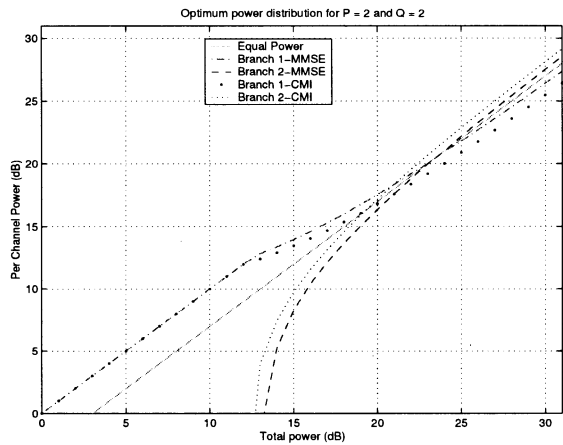
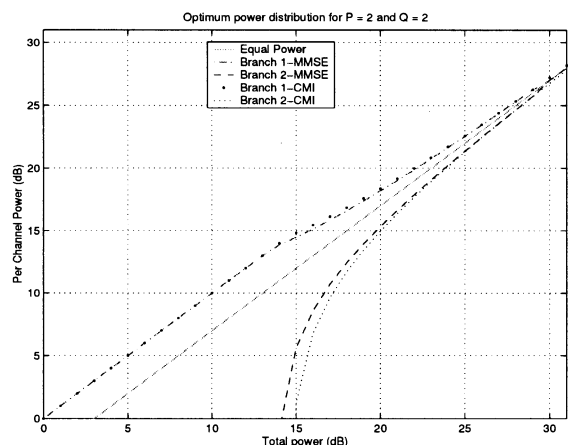
*Remark:* A MIMO channel, for which only the diagonal elements of  $\mathbf{H}_V$  are nonzero (a highly correlated channel), has identical power assignment as the above MISO case.

4) *Transmit and Receive Correlated Channel:* For the channel model in Section II-B, a closed-form solution exists for the special case when either  $\Lambda_T$  or  $\Lambda_R$  is equal to  $\sigma^2 \mathbf{I}$  for some  $\sigma^2$ . In such a case, (35) and (36) reduce to having only one summation, and the resulting closed-form solution is similar to (43).

## V. OPTIMAL SIGNAL DESIGN— INTERPRETATION AND SIMULATIONS

The optimal signal is a block of length  $K \leq P$  and has a diagonal structure given by  $\mathbf{S}_V = \Lambda_S$ . The block  $\mathbf{S}_V$  represents beams transmitted in succession along the fixed virtual transmit angles, with the powers given by the water filling arguments (35) and (36) for the MMSE and CMI criteria, respectively. Basically, the scattering environment is scanned along the virtual transmit angles, one by one, and the presence of scatterers is determined by measuring the signal along the receive virtual angles for each transmitted beam. In other words, the  $i$ th transmitted beam is used to estimate the  $i$ th column of  $\mathbf{H}_V$ . Depending on  $\mathbf{R}_V$  and the SNR, power is assigned to the beams by water filling, which identifies the set of virtual transmit angles that couple strongly enough with receive angles. Hence, the block length  $K$ , which is exactly equal to the size of this set, depends on the SNR and  $\mathbf{R}_V$ . In particular, for low SNR,  $K = 1$ , whereas for high SNR  $K$  has a maximum value equal to the number of active columns determined from  $\mathbf{R}_V$  (which has a maximum of  $P$ ) and for medium SNR,  $1 \leq K \leq P$ . For high SNR, the CMI (MMSE) criterion assigns the power to the transmit angles in proportion to the sum ( $\sqrt{\text{sum}}$ ) of the active elements with which they couple at the receiver. As the SNR decreases further, the weakest transmit beam (as determined by the water filling criteria) is dropped. As the SNR decreases, this process continues until finally, the CMI (MMSE) criterion assigns all the power to the strongest transmit beam, that is, one for which the sum (sum of squares) of the variances of the corresponding virtual receive elements is maximum.



Fig. 6. Optimal power distribution for  $P = 2$  and  $Q = 1$ .Fig. 8. Optimal power distribution for  $P = 2$ ,  $Q = 2$ , and  $\mathbf{R}_V = \mathbf{R}_V^{(3)}$ .Fig. 7. Optimal power distribution for  $P = 2$ ,  $Q = 2$ , and  $\mathbf{R}_V = \mathbf{R}_V^{(2)}$ .Fig. 9. Optimal power distribution for  $P = 2$ ,  $Q = 2$ , and  $\mathbf{R}_V = \mathbf{R}_V^{(4)}$ .

These interpretations of our analytical results are confirmed by the numerical results presented in Figs. 6–9. In all the figures, the total transmit power (in decibels) along the  $x$ -axis is given by  $10 \log_{10}(\beta/\sigma^2)$ , whereas the  $y$ -axis shows the branch powers in decibels given by  $10 \log_{10}(\beta_i/\sigma^2)$ . Powers are plotted for both MMSE and CMI criteria, and the equal power assignment is also plotted for comparison. Fig. 6 shows the power assignments for the MISO case with  $P = 2$ , where the covariance matrix is given by  $\mathbf{R}_V = \text{diag}([1 \ 0.05])$  or, equivalently, for a MIMO case with  $P = Q = 2$  and  $\mathbf{R}_V^{(1)} = \text{diag}([1 \ 0 \ 0 \ 0.05])$ , where the first two components are the variances of the elements in the first column of  $\mathbf{H}_V$ , and the next two are those of the second column. The number of active elements in both the columns is one, i.e.,  $Q_1 = Q_2 = 1$ . Observe that the second transmit beam gets nonzero power only after the SNR increases to about 13 dB (this behavior is similar in the next three figures too). For high SNR, however, both transmit beams get equal power. The following three figures are also for the MIMO case with  $P = Q = 2$ . In Fig. 7, we make the variance of the first element of the second column of  $\mathbf{H}_V$  nonzero to get  $\mathbf{R}_V^{(2)} = \text{diag}([1 \ 0 \ 0.01 \ 0.05])$  so that the number of active elements in the two columns of  $\mathbf{H}_V$  are  $Q_1 = 1$  and  $Q_2 = 2$ . Observe that for high SNR, the second transmit beam gets 66% (58%) power according to the CMI (MMSE) criterion. The powers are reversed in Fig. 8, where  $\mathbf{R}_V^{(3)} = \text{diag}([1 \ 0.1 \ 0 \ 0.05])$  so that now, we have  $Q_1 = 2$  and  $Q_2 = 1$ . Finally, in Fig. 9, we have  $\mathbf{R}_V^{(4)} =$

$\text{diag}([1 \ 0.1 \ 0.01 \ 0.05])$  and  $Q_1 = Q_2 = 2$ . Note that at high SNR, both branches get equal power, as in Fig. 6. However, the power assignment in the medium SNR range is different in the two figures. In all cases, as the SNR decreases the weaker beam is dropped, and the stronger beam gets all the power. The MSE is plotted in Fig. 10 for the first and second cases, corresponding to  $\mathbf{R}_V^{(1)}$  (Fig. 6) and  $\mathbf{R}_V^{(2)}$  (Fig. 7). As expected, the MSE decreases with increasing SNR. Note the change in MSE behavior at approximately 13 dB when the second transmit angle starts getting nonzero power. In case one, both criteria give identical power assignments and, hence, have equal MSE. In case two, however, the MMSE provides lower MSE than the CMI criterion, as expected.

*Remark:* The coherence interval  $T_c$  is determined by the fading rate and is smaller for fast fading compared with slow fading. Consider a fixed fraction of data symbols transmitted in each block so that  $T_d/T_c = \alpha$  for some fixed  $0 < \alpha < 1$ . Since  $T_c = T_d + N_t K$ , we can write  $T_d/T_c = 1 - N_t K/T_c$ . Let the subscript low and high denote low and high SNRs, respectively. Thus, for a constant fraction of data symbols, we have

$$\alpha = \frac{T_{d,\text{low}}}{T_{c,\text{low}}} = \frac{T_{d,\text{high}}}{T_{c,\text{high}}} \implies \frac{T_{c,\text{high}}}{T_{c,\text{low}}} = \frac{N_{t,\text{high}} K_{\text{high}}}{N_{t,\text{low}} K_{\text{low}}}. \quad (44)$$

Now,  $K_{\text{low}} < K_{\text{high}}$ , and since a higher MSE can be tolerated at low SNR, in general,  $N_{t,\text{low}} < N_{t,\text{high}}$ . Then, (44) implies that  $T_{c,\text{low}} < T_{c,\text{high}}$ . Hence, we conclude that from a channel

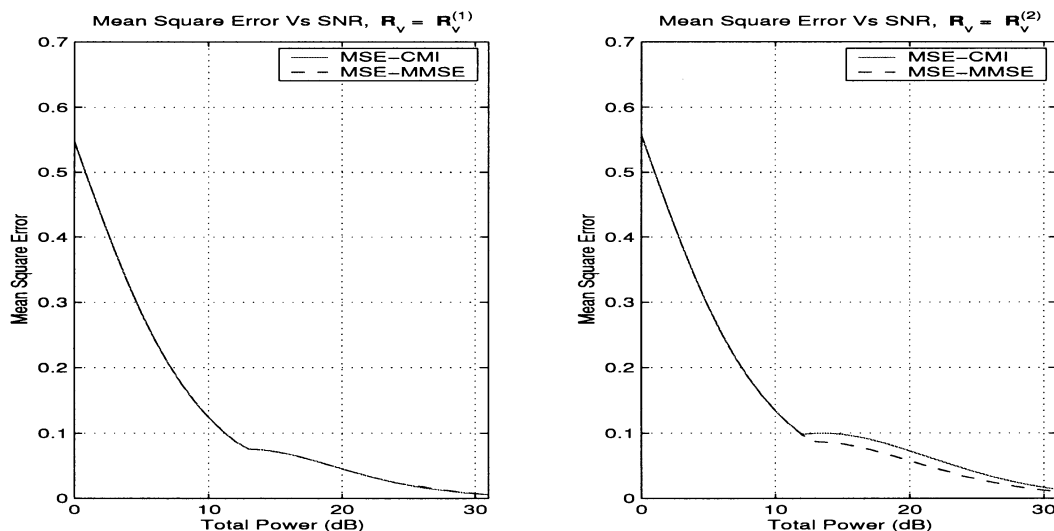


Fig. 10. Mean square error.

estimation viewpoint, for a given fixed fraction of data symbols in each block, faster fading can be tolerated at low SNRs relative to higher SNRs. Note that at lower SNR, the capacity of MIMO systems is lower [6], [13], [23], and a higher bit error rate can be expected. Hence, at lower SNRs, error control coding can be used to attain a desired bit error rate at the expense of transmission rate.

## VI. CONCLUSION

Knowledge of the second-order channel statistics at the transmitter can be efficiently exploited to design the transmit signal for optimal estimation of correlated MIMO channels. Our analysis shows that the optimal transmit signal is a block signal where the block length depends on both the SNR and the channel correlation. The optimal transmit signal is a sequence of beams transmitted along the transmit eigenvectors and sequentially scans the scattering environment to identify the nonzero elements of the virtual channel matrix. The powers of the transmitted beams are determined by the water filling solutions arising from minimum MSE and maximum CMI criteria. Water filling assigns nonzero power only along those directions (beams), where significant scattering is present, as a function of the SNR. Thus, the block length is equal to the number of transmit beams that are assigned nonzero power. For high SNR, all directions (beams) that have significant scatterers get nonzero power, whereas for low SNR, only the direction (beam) with the strongest scatterers gets all the power. For medium SNR, the number of beams with nonzero power is between these two extremes. Consequently, a higher fading rate can be tolerated at lower SNRs relative to higher SNRs. The above observations are consistent with those in [6] and [13], where the capacity analysis of a MIMO system assuming covariance feedback shows that capacity depends on a dominant subset of channel eigenvalues determined by the SNR.

## APPENDIX THEORY OF MAJORIZATION

We introduce the basic necessary concepts of majorization that we require in the derivation of the optimal transmit signal.

For any  $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ , let  $x_{[1]} \geq \dots \geq x_{[n]}$  denote the components of  $\mathbf{x}$  in decreasing order.

*Definition 1:* For vectors  $\mathbf{x}, \mathbf{y} \in \mathcal{A} \subset \mathbb{R}^n$ , vector  $\mathbf{y}$  majorizes  $\mathbf{x}$  on  $\mathcal{A}$  if

$$\sum_{i=1}^k x_{[i]} \leq \sum_{i=1}^k y_{[i]}, \quad k = 1, \dots, n-1$$

$$\sum_{i=1}^n x_{[i]} = \sum_{i=1}^n y_{[i]}.$$

The notation  $\mathbf{x} \prec \mathbf{y}$  means  $\mathbf{x}$  is majorized by  $\mathbf{y}$  on  $\mathcal{A}$ , or  $\mathbf{y}$  majorizes  $\mathbf{x}$  on  $\mathcal{A}$ .

Majorization makes precise the vague notion that the components of a vector  $\mathbf{x}$  are *less spread out* or *more nearly equal* than the components of vector  $\mathbf{y}$ .

*Definition 2:* A real-valued function  $f$  defined on a set  $\mathcal{A} \subset \mathbb{R}^n$  is said to be *Schur-convex* on  $\mathcal{A}$  if

$$\mathbf{x} \prec \mathbf{y} \text{ on } \mathcal{A} \implies f(\mathbf{x}) \leq f(\mathbf{y}). \quad (45)$$

The function  $f$  is *strictly Schur-convex* if the inequality is strict whenever  $\mathbf{x} \prec \mathbf{y}$  but  $\mathbf{x}$  is not a permutation of  $\mathbf{y}$ .  $f$  is *Schur-concave* (respectively, *strictly Schur-concave*) if the inequality in (45) is reversed. It follows that  $f$  is Schur-convex on  $\mathcal{A}$  if and only if  $-f$  is Schur-concave on  $\mathcal{A}$ .

*Definition 3:* A real-valued function  $f$  defined on a set  $\mathcal{A} \subset \mathbb{R}^n$  is said to be *symmetric* on  $\mathcal{A}$  if for every permutation matrix  $\mathbf{P}$  s.t.  $\mathbf{P}\mathbf{x} \in \mathcal{A}$

$$f(\mathbf{P}\mathbf{x}) = f(\mathbf{x}). \quad (46)$$

It can be shown that every Schur-convex or Schur-concave function is symmetric.

Next, we recall a key theorem from [22, Th. 9.B.1], which relates the diagonal elements of a symmetric matrix to its eigenvalues.

*Theorem 3:* If  $\mathbf{d} = (d_1, \dots, d_n)$  and  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)$  are vectors of the diagonal elements and eigenvalues of a symmetric matrix, then  $\mathbf{d} \prec \boldsymbol{\lambda}$ .

In fact, it can be shown that if  $\mathbf{d} \prec \boldsymbol{\lambda}$ , then there exists a real symmetric matrix with diagonal elements  $(d_1, \dots, d_n)$  and eigenvalues  $(\lambda_1, \dots, \lambda_n)$ .

A necessary and sufficient condition for a function  $f$  to be Schur-convex or Schur-concave is given in [22, Th. 3.A.4]. However, for purposes of this discussion, it suffices to recall the following theorem [22, Th. 3.C.1].

*Theorem 4:* If  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  is convex, then the symmetric convex function  $f(\mathbf{x}) = \sum_{i=1}^n \phi(x_i)$  is Schur-convex.

We immediately have the following corollary.

*Corollary 1:* If  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  is concave, then the symmetric concave function  $f(\mathbf{x}) = \sum_{i=1}^n \phi(x_i)$  is Schur-concave.

#### Derivations for High and Low SNR Cases

We provide derivations for the MMSE criterion, whereas the derivations for the CMI criterion can be obtained similarly.

1) *High SNR–MMSE Criterion:* Consider the high RSNR case, where  $\text{RSNR}(i, j) \gg 1$  for the set of active channel coefficients.

From (35), using  $\sigma^2 \ll \mathbf{R}_V[(i-1)Q + j, (i-1)Q + j]|\Lambda_S(i, i)|^2$ , we can write

$$\begin{aligned} \Lambda_M &\approx \arg \min_{\Lambda_S} \sum_{i=1}^P \sum_{j=1}^Q \frac{\sigma^2}{|\Lambda_S(i, i)|^2} \\ &= \arg \min_{\Lambda_S} \sum_{i=1}^P \frac{\sigma^2 Q_i}{|\Lambda_S(i, i)|^2} \end{aligned}$$

subject to the constraint  $\text{tr}(\Lambda_S^H \Lambda_S) \leq \beta$ , where  $Q_i$  is the number of active elements in the  $i$ th column. Optimizing using Lagrange multipliers gives

$$\beta_i = |\Lambda_S(i, i)|^2 = \frac{\sqrt{Q_i} \beta}{\sum_{l=1}^P \sqrt{Q_l}} \quad i = 1, \dots, P. \quad (47)$$

The water filling solution for the CMI criterion can be obtained similarly by noting that  $\log(1+x) \approx \log(x)$  for  $x \gg 1$  to get

$$\beta_i = |\Lambda_S(i, i)|^2 = \frac{Q_i \beta}{\sum_{l=1}^P Q_l} \quad i = 1, \dots, P. \quad (48)$$

2) *Low SNR–MMSE Criterion:* Consider the low RSNR case, where  $\text{RSNR}(i, j) \ll 1, \forall i, j$ .

From (20) and (35), minimization of the MMSE can be written as maximization of the second term in (20), i.e.,

$$\begin{aligned} &\Lambda_M \\ &= \arg \max_{\Lambda_S} \sum_{i=1}^P \sum_{j=1}^Q \frac{\mathbf{R}_V^2[(i-1)Q + j, (i-1)Q + j] |\Lambda_S(i, i)|^2}{\sigma^2 + \mathbf{R}_V[(i-1)Q + j, (i-1)Q + j] |\Lambda_S(i, i)|^2} \\ &\approx \arg \max_{\Lambda_S} \sum_{i=1}^P \sum_{j=1}^Q \frac{\mathbf{R}_V^2[(i-1)Q + j, (i-1)Q + j] |\Lambda_S(i, i)|^2}{\sigma^2} \\ &= \arg \max_{\Lambda_S} \sum_{i=1}^P \left( \sum_{j=1}^Q \mathbf{R}_V^2[(i-1)Q + j, (i-1)Q + j] \right) \cdot \\ &\quad \cdot |\Lambda_S(i, i)|^2 \end{aligned} \quad (50)$$

subject to the constraint  $\text{tr}(\Lambda_S^H \Lambda_S) \leq \beta$ . The last equation is a constrained linear optimization problem, and the solution assigns all the power  $\beta$  to the  $k$ th (virtual) transmit antenna such that

$$k = \arg \max_i \sum_{j=1}^Q \mathbf{R}_V^2[(i-1)Q + j, (i-1)Q + j]. \quad (51)$$

The water filling solution for the CMI criterion can be obtained similarly by noting that  $\log(1+x) \approx x$  for small  $x$ . All the power  $\beta$  to the  $k$ th (virtual) transmit antenna is such that

$$k = \arg \max_i \sum_{j=1}^Q \mathbf{R}_V[(i-1)Q + j, (i-1)Q + j]. \quad (52)$$

#### REFERENCES

- [1] C. N. Chua, J. M. Kahn, and D. Tse, "Capacity of multi-antenna array systems in indoor wireless environment," in *Proc. IEEE Global Telecommun. Conf.*, 1998, pp. 1894–1899.
- [2] E. Telatar, "Capacity of multi-antenna Gaussian channels," *AT&T Bell Labs Intern. Tech. Memo.*, June 1995.
- [3] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.
- [4] A. J. Paulraj and C. B. Papadias, "Space-time processing for wireless communications," in *IEEE Signal Processing Mag.*, Nov. 1997, pp. 49–83.
- [5] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice-Hall PTR, 1998.
- [6] S. A. Jafar, S. Vishwanath, and A. Goldsmith, "Channel capacity and beamforming for multiple transmit and receive antennas with covariance feedback," in *Proc. Int. Conf. Commun.*, 2001, pp. 2266–2270.
- [7] A. M. Sayeed, "Deconstructing multi-antenna fading channels," *IEEE Trans. Signal Processing*, vol. 50, pp. 2563–2579, Oct. 2002.
- [8] D. Shiu, G. Foschini, M. Gans, and J. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Communications*, vol. 48, pp. 502–512, Mar. 2000.
- [9] D. Chizhik, F. Rashid-Farrokhi, J. Ling, and A. Lozano, "Effect of antenna separation on the capacity of BLAST in correlated channels," *IEEE Commun. Lett.*, vol. 4, pp. 337–339, Nov. 2000.
- [10] K. Yu, M. Bengtsson, B. Ottersten, P. Karlsson, and M. Beach, "Second order statistics of NLOS indoor MIMO channels based on 5.2 GHz measurements," in *Proc. IEEE Global Telecommun. Conf.*, 2001.
- [11] G. G. Giannakis and S. Zhou, "Optimal transmit-diversity precoders for random fading channels," in *Proc. Globecom*, 2000, pp. 1839–1843.
- [12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [13] J. H. Kotecha and A. M. Sayeed, "On the capacity of correlated MIMO channels with covariance feedback," in *Proc. Int. Symp. Inform. Theory*, 2003.
- [14] J. H. Kotecha, Z. Hong, and A. M. Sayeed, "Coding and diversity gain tradeoff in space-time codes for correlated MIMO channels," in *Proc. Globecom*, 2003.
- [15] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, vol. CAS-25, pp. 772–781, Sept. 1978.
- [16] A. L. Moustakas, H. U. Baranger, L. Balents, A. M. Sengupta, and S. H. Simon, "Communication through a diffusive medium: Coherence and capacity," *Science*, vol. 287, pp. 287–290, 2000.
- [17] J. R. Fonollosa, R. Gaspa, X. Mestre, A. Pages, M. Heikkilä, J. P. Kermaoal, L. Schumacher, A. Pollard, and J. Ylitalo, "The IST METRA project," *IEEE Commun. Mag.*, vol. 40, pp. 78–86, July 2002.
- [18] H. Ozelik, M. Herdin, H. Hofstetter, and E. Bonek, "A comparison of measured  $8 \times 8$  MIMO systems with a popular stochastic channel model at 5.2 GHz," in *Proc. ICT*, 2003.
- [19] B. Hassibi and B. Hochwald, "Optimal training in space-time systems," in *Conf. Rec. Thirty-Fourth Asilomar Conf. Signals, Syst., Comput.*, vol. 1, 2000, pp. 743–747.
- [20] M. R. Baissas and A. M. Sayeed, "Pilot-based estimation of time-varying multipath channels for coherent CDMA receivers," *IEEE Trans. Signal Processing*, vol. 50, pp. 2037–2049, Aug. 2002.

- [21] J. K. Cavers, "Optimized use of diversity modes in transmitter diversity systems," in *Proc. Veh. Technol. Conf.*, 1999, pp. 1768–1773.
- [22] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Applications*. New York: Academic, 1979.
- [23] "Canonical statistical models for correlated MIMO fading channels and capacity analysis," J. H. Kotecha and A. M. Sayeed, in preparation, preprint: <http://www.ece.wisc.edu/~jkotecha/>.

**Jayesh H. Kotecha** (M'04) received the B.E. degree in electronics and telecommunications from the College of Engineering, Pune, India, in 1995 and the M.S. and Ph.D. degrees from the State University of New York at Stony Brook in 1996 and 2001, respectively, both in electrical and computer engineering.

Since January 2002, he has been with the University of Wisconsin, Madison, as a post-doctoral researcher. His research interests are primarily in communications, signal processing, and information theory. Currently, his focus is in adaptive signal processing and particle filters, sensor networks, and physical layer issues of multiantenna wireless communications, like capacity analysis, transceiver design, channel modeling, and space-time coding.

**Akbar M. Sayeed** (S'89–M'97–SM'02) received the B.S. degree from the University of Wisconsin-Madison, in 1991 and the M.S. and Ph.D. degrees in 1993 and 1996, respectively, from the University of Illinois at Urbana-Champaign, all in electrical and computer engineering.

While at the University of Illinois, he was a research assistant with the Coordinated Science Laboratory and was also the Schlumberger Fellow in signal processing from 1992 to 1995. From 1996 to 1997, he was a postdoctoral fellow at Rice University, Houston, TX. Since August 1997, he has been with the University of Wisconsin-Madison, where he is currently an Associate Professor of electrical and computer engineering. His research interests are in wireless communications, sensor networks, statistical signal processing, wavelets, and time-frequency analysis.

Dr. Sayeed received the National Science Foundation CAREER Award in 1999 and the Office of Naval Research Young Investigator Award in 2001. He served as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS from 1999 to 2002.